COOPERATIVE FILE INTERFACES: HOW SAS® & BMDP® CAN WORK TOGETHER ON VAX™/VMS

Thalene T. Mallus
The National Institute of Mental Health

ABSTRACT

Addressing compatibility problems between software products is common among data managers and analysts. This paper will discuss the flexibility of the SAS System to interface seamlessly with other file formats and statistical packages under the VMS operating system.

Specifically, SAS's ability to interface with BMDP, a library of statistical analysis programs developed at the UCLA Health Sciences Computing Facility, will be examined. Code examples for the use of the BMDP interface library engine, PROC CONVERT, and an overview of the omitted VMS SAS documentation for PROC BMDP will be discussed. Advantages and disadvantages of each interface method will be explored.

INTRODUCTION

As the nation's largest medical research center, the National Institutes of Health are the Federal Government's primary facilities for biomedical investigation. In the pursuit of knowledge to serve the health needs of the nation's people, the NIH encourage collaborative efforts among scientists on the NIH campus, as well as among university investigators and collaborators from other countries.

As with any large institute, the NIH have available to their investigators multiple resources for data management and analysis. The wealth of resources at the NIH coupled with collaborative efforts often results in multiple file formats and requests for analyses from specific statistical packages differing from the original file formats. Such compatibility problems are commonly shared by those who are responsible for managing and analyzing data.

The purpose of this paper is to elaborate on the interface capabilities of the SAS System, and to organize in one source, a discussion of the BMDP interface procedures as they apply to the VMS operating system. Specifically, this paper will elaborate on a) How SAS can interface with BMDP data; b) How BMDP files can be converted to SAS data sets; c) How BMDP procedures can be directly used with SAS data[1].

A DATA SET EXAMPLE

Psychophysiology Data

To illustrate the interface capabilities between SAS and BMDP, a data sample from a NIMH project examining psychophysiological changes in psychiatric patients will be used. This study involved measurements obtained from patients and controls on a variety of physiological and psychological tests during double blind medication phases in the clinical treatment of schizophrenia.

A subset of variables, including demographic measures and sweat gland activity (Galvanic Skin Response, GSR) will be used to illustrate the interface methods under discussion. Table 1 describes the example variables to be utilized, and Table 2 represents a listing of the ASCII data values.

Table 1. Psychophysiology Sample Data

Variables	Description
GROUP	Schizophrenia group code. (2=Males, 52=Females)
SUB	Subject Number. (In ascending order per group)
GENDER	Sex of subject. (1=Males, 2=Females)
HAND	Preferred Hand. (1=Right Handed, 2=Left Handed)
PCBO_GSR	GSR fluctuations per minute during placebo baseline phase.
MED1_GSR	GSR fluctuations per minute during medication #1 baseline

phase.

MED2_GSR GSR fluctuations per minute during medication #2 baseline phase.

Table 2. Psychophysiology Data Values: (SCHIZ.DAT)

Group	Sub	Gender	Hand	Pcbo_ GSR	Med1_ GSR	Med2_ GSR
2	130	1	2	12.8	8.6	1.8
2	136	1	1	1.0		0.6 0.2
2	137	1	1	2.2		1.0 0.8
2	138	1	1	6.6		2.6 0.0
2	140	1	2	2.2	.	0.0
52	109	2	1	.	0.0	0.0
52	116	2	2	0.8	0.0	0.0
52	128	2	1	0.0	1.4	0.0
52	131	2	2	8.2	4.5	0.0
52	135	2	2	2.4	0.0	0.0

Saving the Data as Binary System Files

Since the procedures under discussion involve either the conversion of BMDP files to SAS data sets or the use of BMDP procedures on SAS data, two sets of code will be provided to illustrate how the ASCII data set above could be saved as either a SAS permanent data set or a BMDP save file, respectively[2].

Saving the Data as a SAS Permanent Data Set

The following SAS code produces the VMS SAS data set: SCHIZ.SASEB$DATA. Observations are saved in sorted order by GROUP and SUB.

```
FILENAME in '[vmsdir]schiz.dat';
LIBNAME out '[vmsdir]';
DATA schiz;
INFILE in;
INPUT group sub gender hand
        pcbo_gsr med1_gsr med2_gsr;
PROC SORT DATA=schiz OUT=out.schiz;
        BY group sub;
TITLE 'Creating SAS Dataset:
        SCHIZ.SASEB$DATA';
```

Saving the Data as a BMDP Save File

The following BMDP code produces a similar VMS BMDP binary file, SCHIZ.SAV with an internal BMDP member named EXAMPLE. Observations are again saved in sorted order by GROUP and SUB.

```
/PROBLEM TITLE = 'Creating a BMDP Save File
                  Named: SCHIZ.SAV'.
/INPUT  VARIABLES=7.
        FORMAT=FREE.
        CASES=10.
        RECLEN=89.
        MCHAR='.'.
        SORTS= group, sub.
        FILE='[vmsdir]schiz.dat '.
/VARIABLE NAMES = group, sub, gender, hand,
                  pcbo_gsr, med1_gsr, med2_gsr.
/SAVE   FILE='schiz.sav'.
        CODE=example.
        LABEL='Schiz Data Saved As a BMDP Save File'.
        NEW.
/END
```

INTERFACING WITH OTHER VENDORS

Reading BMDP Save files with SAS

Typically, collaborative efforts in our laboratory necessitate the sharing of data files. Frequently investigators employ preferred statistical packages to manage or analyze data. The following method can be used when a BMDP save file exists on VMS and a request is made to perform SAS procedures on the BMDP binary file.

The BMDP Interface Library Engine

With the introduction of Version 6 of the SAS System, a very useful set of internal routines called engines became available. SAS uses these engines to access a variety of data formats. This concept is known as Multiple Engine Architecture™. (For a complete discussion of SAS's Multiple Engine Architecture™, the reader should refer to the *SAS Companion for the VMS Environment, Version 6.)*

Briefly, there are two types of SAS library engines,

native and *interface*. Native library engines access data sets saved in a variety of SAS formats, such as the default Version 6 format (V608), the older SAS Version 5 format (V5), the SAS transport engine (XPORT), as well as several others.

The Interface library engines are routines that allow SAS to directly read other vendor's file formats. SAS is able to access BMDP, SPSS, and OSIRIS files without prior conversion of the data to an ASCII or portable format. It should be noted that because these engines are sequential, they cannot be used with several of the FSP procedures (FSBROWSE, FSEDIT, FSVIEW) or with the POINT= option of the SET statement.
The syntax for the BMDP interface library engine is as follows:

LIBNAME libref **BMDP** 'filename'**;**

Libref refers to the SAS libref, BMDP is the keyword for the BMDP library engine, and filename refers to the BMDP physical file name.

In the example data, if the psychophysiology data were saved as a VMS BMDP save file (SCHIZ.SAV with a BMDP code= EXAMPLE), it would be very simple to read this BMDP data file directly and utilize SAS procedures. Typically, a PROC CONTENTS is beneficial in order to orient to the data:

```
LIBNAME in BMDP '[vmsdir]schiz.sav';
TITLE  'Reading a BMDP Save
        File with SAS';
TITLE2 'Using PROC CONTENTS
        to Orient to the Data';
PROC CONTENTS DATA=in.example;
```

CONVERTING BMDP FILES

Converting a BMDP Save File to a SAS Permanent Data Set: Two Methods

In order to simplify data management, let's assume a decision has been made to maintain all data in the SAS format. In this situation it would be beneficial to have access to a procedure which efficiently converts a BMDP save file to a SAS permanent data set.

In SAS there are two methods to perform such a conversion: The BMDP Engine with PROC COPY and PROC CONVERT.

The BMDP Engine with PROC COPY

Because the BMDP engine is a read only engine, it can only access a BMDP file for SAS to utilize later in procedure steps. By itself, the BMDP engine cannot convert a BMDP file to a permanent SAS data set. However, if the PROC COPY command is used in conjunction with the BMDP engine, it is possible to convert the BMDP binary file to a SAS data set by simply adding a few additional lines of code. (For a detailed explanation of the PROC COPY command the reader is referred to *SAS Procedures Guide, Version 6*.)

Continuing with the psychophysiology data, the following code would convert the BMDP save file (SCHIZ.SAV, code=EXAMPLE) to a permanent SAS data set (EXAMPLE.SASEB$DATA). See Appendix A, LOG 1 for output results.

```
LIBNAME in BMDP '[vmsdir]schiz.sav';
LIBNAME out '[vmsdir]';
TITLE 'Saving a BMDP Save File as
       a SAS File with PROC COPY';
PROC COPY IN=in OUT=out;
SELECT example;
```

Note that in this method, SAS automatically extracts the BMDP member (EXAMPLE) from the BMDP save file (SCHIZ.SAV) and uses the BMDP member name as the prefix for the new SAS data file (EXAMPLE.SASEB$DATA).

Unlike SAS VMS data sets, BMDP save files can have multiple members associated with the same external filename. In the psychophysiology data example, there is only one member (EXAMPLE) contained in the data set SCHIZ.SAV. However, for BMDP files with multiple members, the PROC COPY statement in conjunction with the BMDP engine would copy *all* members of the BMDP save file to the SAS data library by default.

Further, if only certain members of the BMDP file were to be saved, a SELECT statement could be issued along with PROC COPY to reduce the number of data sets extracted and saved to the SAS data library.

In the above psychophysiology data example, if we imagine that SCHIZ.SAV had multiple BMDP

members, the SELECT statement would specifically convert the named BMDP member (EXAMPLE) to a SAS data set. Since the current BMDP data set contains only one member, the SELECT statement is, in this case, redundant.

PROC CONVERT

Like the BMDP engine with PROC COPY, PROC CONVERT can be used to transform BMDP save files to permanent SAS data sets. PROC CONVERT is an older method which was employed prior to SAS Version 6 and Multiple Engine Architecture™.

The syntax for PROC CONVERT is as follows:

PROC CONVERT product-specification <options>;

select options include:

BMDP=fileref **(CODE**= code **CONTENT**= type**)**

Product-specification refers to the vendor type, which in this case is BMDP. The CODE option allows a specific BMDP member to be chosen, and the CONTENT option allows the type of data in the BMDP save file to be specified.

From the example data, the following program code would similarly convert SCHIZ.SAV (BMDP) to EXAMPLE.SASEB$DATA(SAS). See Appendix A, LOG 2 for output results.

```
FILENAME in '[vmsdir]schiz.sav';
LIBNAME out '[vmsdir]';
PROC CONVERT BMDP=in
    (CODE=example CONTENT=DATA)
     OUT=out.example;
TITLE   'Using PROC CONVERT
         to Transform a BMDP
         Save File';
```

Although PROC CONVERT performs essentially the same transformation as the BMDP engine with PROC COPY, the procedure's limitations should be noted. By default, PROC CONVERT will only access and save the first member of a BMDP save file. Since the CODE= option of PROC CONVERT is only designed to select a single specified BMDP member, this limitation can be problematic if all or multiple members of a BMDP file are to be retained.

Additionally, the reader should also consider that the BMDP engine with PROC COPY will automatically create a permanent SAS data set based on the BMDP member name(s). In PROC CONVERT, if a two-level SAS name is not specified on the OUT= option, the resulting data set will be stored in the SAS WORK data library and will not be permanently saved.

Consequently, although both methods can be employed to convert BMDP files to SAS data sets, the BMDP engine with PROC COPY affords the user much more flexibility.

The Conversion Methods: A Discussion

Both conversion methods can be advantageous in that they have the ability to directly read other types of binary files as if these files were SAS data sets. By directly accessing another vendor's files, redundant data processing can be avoided. Either method will avoid the interim step of converting the data to a portable format, or back to an ASCII file and reread as raw data with lengthy descriptive statements. Data values along with associated variable names, transformed variables, and grouping values can be retained from the BMDP save file and are directly written to a SAS data set.

Another advantage is that through the use of these techniques, data management can be simplified. By converting the BMDP save files to SAS data sets, all data can be managed in the same format, thus, eliminating the need to store multiple copies of the same data in a variety of formats.

Managing the data as permanent SAS data sets has numerous advantages since the SAS System has greater data management capabilities. Unlike BMDP files, SAS data sets can be updated interactively with products like SAS/FSP®, numeric or character variables can be used with SAS procedures, and character variables up to 200 bytes can be represented in the data as compared to BMDP's very limited ability to utilize character data.

Since both methods employ a read only engine, a minor limitation to these procedures is that a separate SAS procedure step is required to save the BMDP file as a permanent SAS data set. Once implemented, these procedures will only produce

the converted data files and SAS LOG. No printed output (ie. SAS LIS file) will result from these conversions.

The reader should also note that an additional BMDP variable, named USE, is created during both conversion methods. Apparently USE is a programming variable which assists BMDP during the conversion process. Once the conversion is completed, the end-user may drop the USE variable, if desired.

Another limitation is that the BMDP engine can only read BMDP save files created under the same operating system. It is not possible, for example, to directly read a BMDP save file originating on an IBM MVS system with SAS on a VMS system. To use the VMS BMDP library engine, the BMDP file must originate on VMS.

Finally, the greatest disadvantage to any of these interface procedures is that new versions of BMDP, or changes in the format of the BMDP save file, may make the interface between BMDP and the current version of SAS incompatible. When system dependent release changes occur, upgrades to the SAS System are supported only when a new version or maintenance release of SAS becomes available[3].

USING BMDP PROGRAMS WITH SAS DATA

Occasionally, the situation arises in which an investigator obtains data in a SAS data set format and wishes to perform BMDP procedures on those data. In our laboratory, this situation is fairly common, either because the investigator is more comfortable with the BMDP code, or has a preference for particular BMDP analysis combinations. To avoid needlessly converting SAS data sets to BMDP save files for a limited number of analyses, SAS has the ability to utilize BMDP subprograms from within a SAS program with a procedure called PROC BMDP.

PROC BMDP

PROC BMDP is a procedure which allows SAS to use BMDP subprograms to analyze SAS data sets.

Although the BMDP procedure (when supported) is usually documented in system dependent manuals, through a publishing oversight, the discussion for PROC BMDP was omitted from the *SAS Companion for the VMS Environment, Version 6.* Readers interested in the complete VMS documentation for this procedure can contact SAS Institute for preliminary SAS draft, *SAS Technical Support Document #TS203.*

The syntax for PROC BMDP is as follows:

PROC BMDP <**options**>**;**
VAR variable list**;**
BY variable list**;**
PARMCARDS;
BMDP Program Code
;

The following options can be used:

- **PROG=BMDP**xx
  For instance, PROG=BMDP1D would invoke the BMDP simple data description program, 1D.

- **DATA=** SAS data set
  Indicates the SAS data set to be processed.

- **INPUTFILE= '**filename**'**
  Specifies the file name that will be defined in the BMDP Input paragraph, FILE= option.

- **CONTENT=DATA|CORR|FREQ|MEAN**
  Advises BMDP whether the SAS data set is a standard SAS data set, a correlation matrix, frequency counts, or variable means.

- **LABEL**= variable (or **LABEL2**=variable)
  Allows BMDP to retain up to two character variables to be defined as case labels. Only the first four characters of the values are used for each label.

- **NOMISS**
  Excludes observations with missing values from a BMDP save file created during the procedure.

The remaining statements for PROC BMDP are as follows:

- **VAR** variable list**;**
  The VAR statement specifies the SAS variables to be passed to the BMDP program. By default, if no VAR statement is

used, all numeric variables are passed to the BMDP program.

- **BY** variable list**;**
  The BY statement may be used with the BMDP procedure to obtain separate analyses on groups defined in previous BY variables. Note that in order to use the BY statement, the data must be sorted previously in order of the BY variables.

- **PARMCARDS;**
  The PARMCARDS statement advises SAS that BMDP program code is to follow.

A Simple PROC BMDP Example

Using the psychophysiology sample data and the SAS data set created earlier (SCHIZ.SASEB$DATA), we will examine a simple PROC BMDP program.
See Appendix B, LIS1 for output results.

```
LIBNAME in '[vmsdir]';
PROC BMDP PROG=BMDP1d DATA=in.schiz
            INPUTFILE='example';
  VAR group sub gender hand
      pcbo_gsr med1_gsr med2_gsr;
PARMCARDS;
/PROBLEM TITLE IS 'Example PROC
                   BMDP Program'.
/INPUT FILE='example'.
       CODE=schiz.
/PRINT METHOD=VAR.
       PAGESIZE=55.
/END
;
```

In the example above, once PROC BMDP is invoked, the SAS data set (SCHIZ.SASEB$DATA) is converted to a format that BMDP understands (EXAMPLE). SAS then utilizes the appropriate BMDP subprogram (1D), BMDP completes the data analysis, and the BMDP output is passed back to the SAS System for display in a SAS LIS file.

Note that the SAS data set name on the DATA= option of the PROC BMDP statement (SCHIZ), and the CODE= option of the BMDP code (SCHIZ), must match. Likewise, the BMDP name specified on the INPUTFILE option of the PROC BMDP statement (EXAMPLE) and the FILE= option within the BMDP subprogram (EXAMPLE) must also match.

Once PROC BMDP has concluded, the user will notice that two VMS DAT files are created by the SAS program.

The first file is EXAMPLE.DAT;1. This is the BMDP file created by the SAS program. The second file is FT15F001.DAT;1 . This is the parmcards file that includes the source code executed by BMDP. Both files are created by SAS for use in the PROC BMDP procedure. If desired, the BMDP file can be retained and read by BMDP in future analyses, and the BMDP source code can be retained for debugging purposes. Because new files are generated with each PROC BMDP session, I have found it useful to automatically delete these files immediately after each run by invoking a VMS command file at the end of the SAS program.

PROC BMDP: A Discussion

As previously discussed, one of the major advantages to the BMDP procedure is that it allows the use of BMDP subprograms directly on a SAS data set. Consequently, requests for BMDP analyses can be completed seamlessly without first converting SAS data sets to BMDP save files.

Because there is no need to create a BMDP file to perform BMDP analyses, the data can be managed in one format, under one system (SAS), which promotes efficiency and streamlines data management.

Another obvious advantage to PROC BMDP is that it allows the user to invoke both SAS and BMDP procedures in the same job. This allows users to choose their favorite programs or the best procedures of either the SAS System or the BMDP subprograms. In a sense, the user has available the best of both analysis worlds.

Finally, the most useful advantage to PROC BMDP is that the procedure allows the user to stack a variety of BMDP subprograms in the same SAS job. Thus, subprograms that would have to be invoked separately with the stand alone BMDP product can be invoked in a single job under SAS. For instance BMDP7D (Histogram program) and BMDP2V (ANOVA program) could be executed in the same SAS session.

The greatest disadvantage to the PROC BMDP procedure is the manner in which character data is handled. BMDP is a software product that is

designed to work with numeric data. BMDP only allows two alphabetic variables of four characters each to be used as variable labels.

If PROC BMDP is to be used with a SAS data set containing numerous (more than two) character variables, the remaining character variables must either be converted to numeric values or PROC BMDP will not pass these character variables to the BMDP subprogram for analysis. The reader should note that the BMDP variable labels must be listed on the PROC BMDP VAR statement.

There are some obvious limitations to using the procedure that warrant mention. In order to use PROC BMDP in SAS, the BMDP program library must be available to the user. This means that in addition to having SAS installed, BMDP Statistical Software must also be installed. At smaller sites, or within departments with budget concerns, the additional licensure of BMDP might be costly or redundant.

Further, any user who intends to employ the PROC BMDP procedure with any regularity will have to become familiar with BMDP syntax and the subprogram analyses. Although most users can utilize their SAS programming skills to assimilate BMDP rather quickly, there might be some individuals who feel that to learn an additional programming language is a burden.

CONCLUSION

In today's research environments, it is commonplace to receive data from collaborative sources in different file formats. Individual researchers or departments often develop preferences for specific statistical packages, analyses, and their associated output.

As new technologies emerge and we are able to receive data in different formats, our expectations for seamless integration of file interfaces have increased. SAS has kept up with the demand for this integration through the use of its' Multiple Engine Architecture™, which allows different vendor's formats to be read; and procedures such as PROC BMDP which have the ability to utilize BMDP subprograms with SAS data sets.

Although the interface methods described in this paper have some limitations, I have found that through their use, data management has been streamlined and analyses from different vendors can be completed with ease and efficiency.

TRADEMARKS

NOTES TO USERS

1. This paper assumes the reader has general knowledge regarding the VMS environment, SAS and BMDP programming conventions.
2. Programs for this paper were run on a MICROVAX Model 3100-40 under VMS 5.5-2 operating system, SAS Release 6.08 TS420, BMDP Release 7.0 (1993).
3. VMS SAS 6.08 and VMS BMDP releases 1990,1993 are compatible if invoked under SAS Maintenance release TS410 or later.

REFERENCES

BMDP Statistical Software, Inc. (March 1993). *BMDP VAX/VMS Installation and Batch User's Guide, Release 7*.

Dixon, W.J. (Ed.) (1992). *BMDP Statistical Software Manual ,Volume 1, Version 7.0*. Berkeley: University of California Press.

SAS Institute Inc. (1992) *Master Index to SAS System Documentation, Version 6*. Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1990) *SAS Companion for the MVS Environment, Version 6*. Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1990) *SAS Companion for the VMS Environment, Version 6*. Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1989) *SAS /FSP Software: Usage and Reference, Version 6*. Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1990) *SAS Language: Reference, Version 6*. Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1990) *SAS Procedures Guide, Version 6*. Cary, NC: SAS Institute Inc.

SAS Institute Inc. *SAS Technical Support Document #TS203*

ACKNOWLEDGEMENTS

Loring Ingraham, Ph.D./ Psychologist/ NIH/NIMH
Theodore Zahn, Ph.D. / Psychologist/ NIH/NIMH
John Ingeholm, Ph.D./ Biomedical Engineer /NIH/NIMH
Elvira Agron/ Statistical Support/ NIH/DCRT
David Shinn / Technical Support/ SAS Institute Inc.
Tom Hahl / Technical Support/ SAS Institute Inc.
Christine Peng / Technical Support / BMDP Statistical Software Inc.

The author may be contacted at:

The Laboratory of Psychology and
Psychopathology, NIMH
Building 10, Room 4C110
10 CENTER DR MSC 1366
BETHESDA MD 20892-1366

****

Phone: (301) 496-7672 ext. 35
FAX: (301) 402-0921, (301) 480-2251
E-Mail :Thalene_Mallus@nih.gov